



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

24.08.2004

REC'D 01 OCT 2004

WIPO

PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03425582.8

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Best Available Copy

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk



Anmeldung Nr:
Application no.: 03425582.8
Demande no:

Anmeldetag:
Date of filing: 09.09.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Semeion
Viale di Val Fiorita 88
00144 Roma
ITALIE

Bezeichnung der Erfindung/~~Title of the invention~~/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

An artificial neural network

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G06N3/00

Am Anmeldetag benannte Vertragsstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

SEMEION

5 Rete neurale artificiale

10 L'invenzione ha per oggetto una rete neurale artificiale comprendente diversi nodi che formano almeno due strati uno dei quali è un cosiddetto strato di input e il secondo dei quali è un cosiddetto strato di output, i nodi dello strato di input formando degli
15 input per l'immissione di dati di input di un database e i nodi dello strato di output formando dei canali di output per fornire i dati di output che sono i risultati dell'elaborazione dei dati di input, essendo previste delle connessioni tra i nodi dello strato di input e i nodi dello strato di output, ciascun nodo dello strato di output eseguendo una trasformazione dei
20 dati di input ricevuti dallo strato di input in dati di output la quale trasformazione comprende una prima fase di trasformazione consistente essenzialmente nel
25 sommare i dati di input ricevuti dai nodi di input ai detti nodi di output pesando i detti dati di input e una seconda fase di trasformazione che trasforma, non linearmente, i risultati ottenuti mediante la prima fase di trasformazione, i dati di output ottenuti me-

diante detta trasformazione eseguita in un nodo di output essendo i dati di output.

Le reti neurali artificiali sono largamente utilizzate per diversi scopi. Generalmente parlando, le
5 reti neurali artificiali appartengono alla famiglia dei cosiddetti algoritmi predittivi che sono in grado di apprendere dai dati di un database al fine di raggiungere un certo livello iniziale di esperienza per la valutazione di situazioni che possono essere de-
10 -scritte mediante i record di dati di un database acquisiti in molti modi differenti. In questo modo un algoritmo adattativo addestrato come le reti neurali artificiali può essere impiegato per valutare le conseguenze di certe situazioni definite da certe caratteristiche che sono rappresentate e codificate dai da-
15 ti o dalle variabili dei record di dati di un database e sulla base dei risultati di questa valutazione una macchina funzionante mediante questi algoritmi adattativi può prendere delle decisioni circa l'applicazione
20 di certe misure o azioni senza la necessità dell'intervento di personale umano.

Gli algoritmi adattativi non smettono di apprendere e possono accumulare e perfezionare la loro esperienza migliorando la loro conoscenza ogni volta che
25 vengono utilizzati, in questo modo aggiungendo, mediante il loro utilizzo, un'ulteriore conoscenza a quella acquisita durante la prima fase di apprendimento.

Questo tipo di algoritmi sono le basi per miglio-

rare le capacità di eseguire automaticamente dei compiti senza l'aiuto o l'intervanto dell'intelligenza umana. Tipicamente essi sono applicati nei robot o in sistemi altamente intelligenti per l'automatizzazione e permettono di espandere l'utilizzo delle macchine con elevati livelli di attendibilità anche nell'esecuzione di compiti che non possono essere definiti in termini di dati rigidi basati su parametri tecnici, fisici, chimici o simili.

Da questo punto di vista tale tipo di algoritmi hanno un carattere tecnico ed un risultato tecnico, poiché lo scopo di questi algoritmi è di provvedere a dispositivi tecnici, come computer che calcolano i dati in un modo simile al modo con cui questi dati sarebbero stati trattati dal cervello umano provvedendo, in un modo più semplice, ad una struttura che è conforme alla nota struttura del cervello.

Gli algoritmi sono utilizzati per risolvere molti problemi tecnici o per fornire risultati tecnici come per esempio per metodi di riconoscimento di immagini o di rilevamento dei contorni.

Effettivamente sono noti diversi tipi di architetture per tali algoritmi adattativi e particolarmente per tali reti neurali artificiali.

Il principio e la struttura generale di tali reti neurali artificiali sono descritti in quanto segue mediante le figure 1 e 3.

In una classica rete ciascun nodo funziona come un elemento che riceve gli input pesati dai nodi di

input, li somma e trasforma il risultato mediante una
funzione non lineare. I nodi sono elementi funzionali
essendo in grado di eseguire trasformazioni dei dati
di input forniti dal database e o da altri nodi e di
5 fornire come output, i dati di output che devono esse-
re inviati agli altri nodi o i dati di output della
rete neurale artificiale.

La fig. 3 chiarisce il meccanismo. Nelle reti
classiche ciascun nodo esegue una trasformazione non
10 lineare su una trasformazione lineare dei suoi input:

$$x_j^{[s]} = F\left(L\left(w_{ji}^{[s]}, x_i^{[s-1]}\right)\right) \quad (1)$$

dove la trasformazione non lineare $F(\cdot)$ è una
15 funzione come, per esempio, una sigmoidale e la tra-
sformazione lineare $L(\cdot)$ è la somma pesata degli in-
put:

$$L\left(w_{ji}^{[s]}, x_i^{[s-1]}\right) = \sum_{i=0}^n w_{ji}^{[s]} \cdot x_i^{[s-1]} \quad (2)$$

Come risultato da (1) e (2):

20

$$x_j^{[s]} = F\left(\sum_{i=0}^n w_{ji}^{[s]} \cdot x_i^{[s-1]}\right) \quad (3)$$

dove

$[s]$: lo strato generico della rete, con $s=1$ per lo

strato di input e valori crescenti per lo strato nascosto e di output;

$x_j^{[s]}$: la variabile di output del nodo j -th nello strato $[s]$;

5 $x_i^{[s-1]}$: l'input i -th del nodo generico nello strato $[s]$ dal nodo i -th nello strato $[s-1]$;

$x_0^{[s-1]}$: un input "falso" al nodo generico nello strato $[s]$, introdotta artificialmente per rappresentare, in
un modo matematicamente conveniente, un utile valore
10 di soglia. Il suo valore è solitamente fissato a 1.

$w_{ji}^{[s]}$: il peso sulla connessione che unisce il nodo i -th
nello strato $[s-1]$ al nodo j -th nello strato $[s]$;
" il numero degli input al nodo.

Inoltre è ben noto che le reti neurali artificia-
15 li possono avere più di due strati, vale a dire lo strato di input e lo strato di output i quali strati sono previsti intermedi agli strati di input e di output e sono chiamati strati nascosti. Una rete neurale artificiale può avere più di uno strato nascosto.
20 Ciascun nodo di ciascuno strato può essere connesso mediante un canale di comunicazione con uno o più o con ciascun nodo dello strato successivo che può essere uno strato nascosto o lo strato di output. Ciascun nodo esegue le trasformazioni sopra menzionate
25 sull'input ricevuto dagli altri nodi dello strato precedente e fornisce ai nodi dello strato successivo un valore di output che è calcolato mediante la trasfor-

mazione sopra menzionata dell'input ricevuto dai nodi dello strato precedente. Sono noti e utilizzati diversi tipi di architetture della connessione tra i nodi del differente o dello stesso strato e la presente invenzione non è limitata a nessuna delle dette note architetture poiché qualsivoglia delle note architetture delle reti neurali possono essere modificate secondo gli insegnamenti della presente invenzione.

Per le reti neurali artificiali è uno scopo importante quello di migliorare la loro attendibilità e la loro capacità di convergere rapidamente ad una soluzione altamente precisa del problema trattato, cioè ad una valutazione altamente precisa della situazione illustrata dai dati di input e quindi ad una predizione o ad un dato di output altamente attendibile.

Le funzioni descritte sopra comprendono dei parametri che sono normalmente impostati casualmente a certi valori di inizializzazione e la fase di apprendimento dell'algoritmo consiste nel fornire all'algoritmo noti dati di input e i corrispondenti noti dati di output. Durante questa fase di apprendimento i parametri, solitamente i pesi applicati ai dati di input ricevuti mediante la prima fase di trasformazione, vengono corretti al fine di ottenere una migliore corrispondenza dei dati di input con i corrispondenti dati di output noti dai record dell'insieme di dati di apprendimento.

Sono noti differenti gruppi di funzioni di apprendimento per far apprendere l'algoritmo. Questi

gruppi di funzioni di apprendimento sono una caratteristica critica dell'algoritmo poiché la fase di apprendimento influenza l'attendibilità computazionale e la velocità computazionale dell'algoritmo e quindi la funzionalità dell'algoritmo. I pesi sono definiti mediante la riduzione della distanza tra gli output computati e i noti output per lo stesso dato di input. A seconda della struttura dei dati spesso durante la minimizzazione dell'errore l'algoritmo rileva solo i minimi locali e non i minimi assoluti in modo tale per cui esso resta attaccato a questi minimi locali senza essere in grado di rilevare i minimi assoluti coincidenti con l'algoritmo meglio addestrato e appreso.

Per una descrizione completa di differenti tipi di reti neurali artificiali si veda per esempio Arbib M. A., (Ed.) (1995a), *The Handbook of Brain Theory and Neural Networks*, A Bradford Book, The MIT Press, Cambridge, Massachusetts, Londra, Inghilterra, 1995.

Buscema M. & Semeion Group (1999a), *Reti Neurali Artificiali e Sistemi Sociali Complessi*, Vol. I: Teoria e Modelli, Franco Angeli, Milano, 1999.

L'invenzione ha lo scopo di migliorare una rete neurale artificiale in modo tale da avere una convergenza migliore, più attendibile e più veloce dell'algoritmo verso la soluzione migliore e di evitare anche i minimi locali durante la fase di apprendimento.

L'invenzione consegue gli scopi sopra menzionati mediante una rete neurale artificiale come descritta

all'inizio comprendente ulteriormente la caratteristica che in ciascun nodo di output la prima fase di trasformazione è una funzione di trasformazione non lineare dei dati di input ricevuti mediante i nodi di output dai nodi di input.

Generalmente una rete neurale artificiale comprende uno strato di input avente un numero predeterminato di nodi di input e uno strato di output avente un numero predeterminato di nodi di output, e inoltre tra lo strato di input e quello di output è previsto almeno uno strato di nodi, il cosiddetto strato nascosto, o più di uno strato nascosto. I nodi di questo strato nascosto sono connessi mediante connessioni pesate ai nodi di input dello strato di input e ai nodi di un ulteriore strato nascosto quando è previsto più di uno strato nascosto o ai nodi di output dello strato di output se è previsto solo uno strato nascosto. Ciascun nodo dell' almeno uno strato nascosto o di più di uno degli strati nascosti e i nodi dello strato di output eseguono una trasformazione dei dati di input ricevuti dallo strato di input o da un precedente strato nascosto in dati di output. Analogamente al caso in cui sono previsti solo due strati questa trasformazione comprende una prima fase di trasformazione consistente nel sommare i dati di input ricevuti dai nodi di input ai detti nodi di output pesando i detti dati di input e una seconda fase di trasformazione che trasforma in modo non lineare i risultati ottenuti dalla prima fase di trasformazione. Quando i nodi sono

i nodi di output dello strato di output i dati di output ottenuti da questa trasformazione eseguita in detti nodi sono i dati di output dell'algoritmo. Quando i nodi sono uno strato nascosto, i dati di output
5 ottenuti da questa trasformazione eseguita in detti nodi sono i dati di input forniti dai nodi di un successivo strato nascosto o ai nodi di output dello strato di output. In ciascun nodo di ciascun strato nascosto e nei nodi di output dello strato di output
10 la prima fase di trasformazione è una funzione di trasformazione non lineare dei dati di input ricevuti mediante i nodi di output o mediante i nodi di uno strato nascosto dai nodi di input dello strato di input o mediante i nodi del precedente strato nascosto.

15 I dati di input dei nodi dello strato di input consistono nei dati di input del database, mentre i dati di output dei nodi dello strato di input sono forniti ai nodi dello strato di output come dati di input e i dati di output dello strato di output consistono nel risultato di elaborazione della rete neurale
20 artificiale.

La prima trasformazione non lineare dei dati di input eseguita da un nodo comprende una prima sotto-fase di trasformazione dei dati di input mediante una
25 funzione sinusoidale e una seconda sotto-fase di trasformazione consistente nella somma dei dati di input dopo la trasformazione mediante la funzione sinusoidale, cioè dopo avere eseguito la prima sotto-fase di trasformazione.

Ciascun nodo almeno dell'almeno uno strato nascosto e dello strato di output può essere descritto come comprendente diversi canali di input per differenti dati di input;

5 ciascun canale essendo associato

un'unità di ricezione per l'esecuzione della prima sotto-fase di trasformazione non lineare della prima fase di trasformazione;

10 un'unità di sommatoria per l'esecuzione della seconda sotto-fase di trasformazione della prima fase di trasformazione sommando i dati di input di ciascun canale trasformati non linearmente ad un valore;

15 e un'unità di trasformazione non lineare per l'esecuzione della seconda fase di trasformazione mediante un filtraggio non lineare del valore ottenuto mediante la prima fase di trasformazione e per fornire il valore di output del nodo che è il valore di input dei nodi di un successivo strato nascosto o dello strato di output. I dati di input possono ulteriormente consistere in un numero predeterminato di variabili
20 in uno spazio di variabili dei dati di input definendo le coordinate nello spazio dei dati di input e ciascuna coordinata nello spazio dei dati di input è trasformata non linearmente nella prima fase di trasformazione
25 in un corrispondente valore di una variabile che viene reso dipendente dalla posizione nello spazio del valore della coordinata rispetto un'onda dello spazio di lunghezza d'onda data, questa dipendenza consistente nel moltiplicare i valori delle coordinate

di input per la lunghezza d'onda di un'onda sinusoidale che vengono poi trasformate nello stesso valore, la lunghezza d'onda su ciascuna coordinata di input essendo regolata durante la fase di apprendimento.

- 5 La trasformazione dei dati di input eseguita da ciascun nodo viene definita dalla seguente equazione:

$$x_j^{[s]} = F\left(G\left(w_{ji}^{[s]}, x_i^{[s-1]}\right)\right) \quad (4)$$

- 10 dove la trasformazione non lineare $F(\cdot)$ è la funzione di filtraggio non lineare della seconda fase di trasformazione e $G(\cdot)$ è la funzione di trasformazione non lineare della prima fase di trasformazione consistente nella somma degli input pesati elaborati in modo non
15 monotono, sinusoidale secondo la seguente funzione:

$$G\left(w_{ji}^{[s]}, x_i^{[s-1]}\right) = \sum_{i=0}^n \sin\left(w_{ji}^{[s]} \cdot x_i^{[s-1]}\right) \quad (5)$$

dove

- $[s]$: lo strato generico della rete, con $s=1$ per lo strato di input e valori crescenti per gli strati nascosti e di output;

$x_j^{[s]}$: la variabile di output del nodo j -th nello strato $[s]$;

$x_i^{[s-1]}$: l'input i -th del nodo generico nello strato $[s]$ dal nodo i -th nello strato $[s-1]$;

$x_0^{[s-1]}$: un input "falso" al nodo generico nello strato $[s]$, introdotto artificialmente per rappresentare, in
5 un modo matematicamente conveniente, un utile valore di soglia che solitamente è fissato a 1.

$w_{ji}^{[s]}$: il peso sulla connessione che unisce il nodo i -th nello strato $[s-1]$ al nodo j -th nello strato $[s]$;

10 n : il numero di input al nodo.

Ciascun nodo esegue una trasformazione dei dati di input secondo la seguente funzione:

$$x_j^{[s]} = F \left(\sum_{i=0}^n \sin(w_{ji}^{[s]} \cdot x_i^{[s-1]}) \right) \quad (6)$$

15 La funzione seno introducendo un processo qualitativo poichè ciascun peso $w_{ji}^{[s]}$ agisce come un parametro 2π / lunghezza d'onda nella coordinata i -th dello spazio degli input del nodo j -th dello strato s -th.

La seconda fase di trasformazione non lineare può
20 essere eseguita mediante una funzione sigmoidale.

Anche i nodi di input dello strato di input pos-

sono eseguire una prima fase di trasformazione non lineare e una seconda fase di trasformazione non lineare dei dati di input ricevuti essendo previste le seguenti unità:

5 un canale di ricezione per i dati di input

un'unità di ricezione associata a detto canale di ricezione per eseguire la prima sotto-fase di trasformazione non lineare della prima fase di trasformazione come definito nelle precedenti affermazioni;

10 un'unità di sommatoria per eseguire la seconda sotto-fase di trasformazione della prima fase di trasformazione sommando i dati di input trasformati non linearmente di ciascun canale ad un valore come definito nelle precedenti affermazioni;

15 e un'unità di trasformazione non lineare per eseguire la seconda fase di trasformazione mediante un filtraggio non lineare del valore ottenuto mediante la prima fase di trasformazione come definito dalla fase precedente e per fornire il valore di output dei nodi di input che è il valore di input dei nodi di un successivo strato nascosto o dello strato di output.

20 La presenza di una specifica doppia relazione non lineare sulle connessioni tra i nodi presenta delle evidenti conseguenze sulle proprietà della rete secondo la presente invenzione sia sulla funzione calcolata e sia sul comportamento di questa rete durante la fase di apprendimento.

L'elaborazione secondo la presente invenzione può essere applicata alle tipologie di reti esistenti come

una modificazione fondamentale delle loro equazioni di apprendimento rappresentante una nuova legge generale di apprendimento. La nuova legge di apprendimento dimostra una considerevole convergenza ed elevate capacità di estrapolazione su basi complesse di dati.

Ulteriori perfezionamenti della presente invenzione sono descritti nelle sotto rivendicazioni.

Le caratteristiche dell'algoritmo secondo la presente invenzione e i vantaggi ottenuti con esso verranno descritti più chiaramente mediante degli esempi dettagliati e mediante i disegni allegati in cui:

La figura 1 illustra una vista schematica dell'elaborazione concettuale nelle reti classiche.

La figura 2 illustra in un modo analogo alla fig. 1 l'elaborazione concettuale secondo le reti della presente invenzione.

La figura 3 illustra schematicamente le variabili dei nodi per il nodo j -th.

La figura 4 illustra schematicamente la struttura di una rete multi strato secondo l'invenzione.

Le figure 5 a 7 illustrano uno schema della soluzione del problema XOR mediante una classica rete neurale Back propagation, di una cosiddetta rete RBF e mediante la rete neurale secondo la presente invenzione.

La figura 8 illustra uno schema del cosiddetto problema della spirale.

Le figure 9 e 10 illustrano degli schemi rappresentanti le soluzioni del problema delle due spirali

ottenute rispettivamente con la rete neurale secondo la presente invenzione e con una classica rete neurale back propagation.

5 Le figure 11a a 11d illustrano degli schemi che dimostrano soluzioni differenti del problema delle due spirali ottenute mediante la rete neurale secondo la presente invenzione con intervalli differenti.

10 La figura 12 illustra una schema della soluzione del problema delle due spirali ottenuta mediante la rete secondo la presente invenzione su 192 punti di input.

La figura 13 illustra gli schemi della funzione di trasferimento dei pesi dai nodi di input ai nodi nascosti in una rete secondo la presente invenzione.

15 La figura 14 illustra lo schema della funzione di trasferimento dei pesi dai nodi nascosti ai nodi di output in una rete secondo la presente invenzione.

20 La figura 15 illustra ulteriori schemi della funzione di trasferimento dei pesi dai nodi di input ai nodi nascosti in una rete secondo la presente invenzione.

25 La figura 16 illustra ulteriori schemi della funzione di trasferimento dei pesi dai nodi nascosti ai nodi di output in una rete secondo la presente invenzione.

La figura 17 illustra uno schema che descrive il confronto tra una classica rete neurale back propagation e la rete neurale secondo la presente invenzione durante il primo esperimento, relativamente alla cor-

rettezza del numero delle unità nascoste.

La figura 18 illustra uno schema rappresentante il risultato del confronto tra una classica rete neurale Back propagation e la rete neurale secondo la
5 presente invenzione sull'insieme di dati del Cancro al Seno.

La figura 19 illustra uno schema rappresentante il risultato del confronto tra una classica rete neurale Back propagation e la rete neurale secondo la
10 presente invenzione sull'insieme dei dati del Cancro al Seno utilizzando l'" Early Stopping ".

La figura 20 illustra uno schema rappresentante il confronto tra una classica rete neurale Back propagation e la rete neurale secondo la presente invenzione
15 ne durante il primo esperimento, relativamente alla correttezza del numero delle unità nascoste con riferimento all'esempio riferito all'insieme di dati sulla Valutazione del Credito Australiano.

La figura 21 illustra uno schema rappresentante il risultato del confronto tra una classica rete neurale Back propagation e la rete neurale secondo la
20 presente invenzione sull'insieme di dati sulla Valutazione del Credito Australiano con "Early Stopping" (Ad-Vr-Pr) e senza (Ad-Pr).

25 L'esempio illustrato nelle figure è utilizzato per le prove si riferisce ad una rete neurale multi-strato avente più di uno strato nascosto interposto tra lo strato di input e lo strato di output (fig.4). I dati del database sono forniti ai nodi dello strato

di input il quale ha dei canali per ricevere questi dati di input. Ciascun nodo dello strato di input è connesso a ciascuno dei nodi di un primo strato nascosto. Ciascun nodo del primo strato nascosto è connesso
5 a ciascun nodo di un ulteriore strato nascosto e così via fino ad un ultimo strato nascosto, ciascun nodo del quale è collegato a ciascuno dei nodi dello strato di output.

Come apparirà più chiaramente nella seguente descrizione teorica e dalle figure 1 a 3, i nodi di una
10 rete neurale sono unità di elaborazione che ricevono dati di input in corrispondenza di uno, due, tre o più canali di input e che eseguono una trasformazione dei dati di input secondo una funzione non lineare per
15 fornire ad un canale di output un valore che è il dato di output calcolato dalla rete neurale se i nodi sono quelli dello strato di output o i quali valori di output vengono forniti come dati di input ai nodi di uno strato successivo come uno strato nascosto o come uno
20 strato di output.

Tipicamente la trasformazione dei dati di input ricevuti da un nodo viene eseguita secondo due fasi. Una prima fase di trasformazione per calcolare l'input netto al nodo, che è normalmente una sorta di somma
25 media o pesata dei dati di input ricevuti tramite un nodo e una seconda fase di trasformazione che è un filtraggio non lineare dei dati di input trasformati secondo la prima fase.

La figura 1 illustra un tipico e noto nodo secon-

do lo stato dell'arte. Le singole variabili delle funzioni sono spiegate nella figura 3.

Secondo la figura 3, per un tipico nodo in un certo strato di una rete neurale artificiale sono utilizzate le seguenti variabili il cui significato è spiegato in quanto segue:

$[s]$ è definito come il numero dello strato generico della rete, con $s=1$ per lo strato di input e con valori crescenti per gli strati nascosti e di output;

10 $x_j^{[s]}$ indica la variabile di output del nodo j -th nello strato $[s]$;

$x_i^{[s-1]}$ indica l'input i -th al nodo generico nello strato $[s]$ dal nodo i -th nello strato $[s-1]$;

$x_0^{[s-1]}$ indica un input "falso" al nodo generico nello strato $[s]$, introdotto artificialmente per rappresentare, in un modo matematicamente conveniente, un utile valore di soglia. Il suo valore è solitamente fissato a 1.

$w_{ji}^{[s]}$ indica il peso sulla connessione che unisce il nodo i -th nello strato $[s-1]$ al nodo j -th nello strato $[s]$;
20 n indica il numero degli input al nodo.

In una rete classica ciascun nodo lavora come un elemento che riceve gli input pesati dai nodi di input, li somma e trasforma il risultato mediante una
25 funzione non lineare. In questo modo nelle reti classiche ciascun nodo effettua una trasformazione non lineare su una trasformazione lineare dei suoi input se-

condo la seguente funzione:

$$x_j^{[s]} = F\left(L\left(w_{ji}^{[s]}, x_i^{[s-1]}\right)\right) \quad (1)$$

dove la trasformazione non lineare $F(\cdot)$ è una funzione
5 come, per esempio, una sigmoidale e la trasformazione
lineare $L(\cdot)$ è la somma pesata degli input:

$$L\left(w_{ji}^{[s]}, x_i^{[s-1]}\right) = \sum_{i=0}^n w_{ji}^{[s]} \cdot x_i^{[s-1]} \quad (2)$$

Come risultato da (1) e (2) viene ottenuta la
classica funzione di trasformazione di un nodo generi-
10 co in una classica rete neurale:

$$x_j^{[s]} = F\left(\sum_{i=0}^n w_{ji}^{[s]} \cdot x_i^{[s-1]}\right) \quad (3)$$

La figura 2 illustra schematicamente la struttura
di un nodo generico in una rete neurale artificiale
secondo la presente invenzione.

15 R indica le unità di ricezione funzionali che
svolgono una prima sotto-fase di trasformazione degli
input della prima fase di trasformazione su ciascun
canale di input del nodo. La prima sotto-fase di tra-
sformazione degli input è una trasformazione non line-
20 are che verrà spiegata in maggiore dettaglio nella se-
guente descrizione.

E indica la seconda sotto-fase di trasformazione
degli input della prima fase di trasformazione che e-
segue la somma dei valori degli input pesati che sono

stati precedentemente trasformati da una funzione non lineare nelle unità di ricezione R.

5 X_i a X_n indicano i valori di input che sono ricevuti mediante il nodo da altri nodi i a n se il nodo appartiene ad uno strato nascosto o allo strato di output o che sono i dati di input del database se il nodo appartiene allo strato di input.

10 w_i a w_n indica il peso applicato a ciascuna connessione lungo la quale il nodo riceve i valori di input dagli altri nodi.

F indica un'ulteriore unità del nodo in cui viene eseguita una fase di trasformazione non lineare sui valori di input che sono stati sottoposti alla prima fase di trasformazione e fornisce in questo modo un
15 valore di output del nodo indicato con Y.

Essendo il nodo illustrato un nodo dello strato di input o di uno strato nascosto il valore di output viene trasmesso attraverso la connessione pesata a ciascuno o ad un certo numero di nodi di uno strato
20 successivo che può essere uno strato nascosto o lo strato di output mentre se il nodo illustrato è il nodo dello strato di output, Y indica il contributo del nodo ai dati di output calcolati dalla rete.

25 L'idea base nell'elaborazione dei dati secondo la rete neurale artificiale della presente invenzione è di fornire ciascun nodo con le unità di ricezione R interposte tra ciascun input e l'unità di sommatoria E. Le unità di ricezione R trasformano appropriatamente in un modo non lineare l'input ricevuto mediante un

nodo da ciascun nodo di input, prima di sommare i contributi degli input in un valore che deve essere trasformato tramite una funzione non lineare, che è la funzione applicata dall'unità F della seconda fase di trasformazione. Il significato delle unità di ricezione R è nell'introduzione di un processo quantitat-
5 qualitativo sul valore di input, in sostituzione ad un semplice processo quantitativo su esso, in analogia a quanto viene fatto negli organismi biologici mediante
10 porte chimiche rispetto a porte a potenziale. Gli aspetti qualitativi della trasformazione vengono ottenuti utilizzando delle funzioni sinusoidali. Per ciascuna coordinata i-th dello spazio degli input, questo permette l'introduzione di una dipendenza di ciascun
15 valore i-th trasformato mediante la posizione nello spazio del valore della coordinata rispetto all'onda dello spazio di lunghezza d'onda data. I valori delle coordinate di input, moltiplicati per la lunghezza d'onda, sono poi trasformati nello stesso valore. La
20 lunghezza d'onda di ciascun ricettore di input viene regolata durante la fase di apprendimento della rete neurale artificiale come risulterà evidente nella seguente descrizione di una forma esecutiva.

Nella rete neurale artificiale seconda la presente invenzione ciascun nodo effettua una trasformazione
25 non lineare su una trasformazione non lineare dei suoi input. La funzione che descrive questa trasformazione è indicata sotto considerando la terminologia delle variabili definita sopra con riferimento alla figura

3:

$$x_j^{[s]} = F\left(G\left(w_{ji}^{[s]}, x_i^{[s-1]}\right)\right) \quad (4)$$

In questo caso la trasformazione non lineare $F(\cdot)$
5 è nuovamente una funzione come, per esempio, una sig-
moidale e $G(\cdot)$ è anche una trasformazione non lineare
che può essere definita come la somma degli input pe-
sati elaborati in modo non monotono, sinusoidale:

$$G\left(w_{ji}^{[s]}, x_i^{[s-1]}\right) = \sum_{i=0}^n \sin\left(w_{ji}^{[s]} \cdot x_i^{[s-1]}\right) \quad (5)$$

10

Come risultato dalla (4) e (5) la trasformazione
effettuata sull'input mediante un nodo in una rete
neurale artificiale secondo la presente invenzione è
descritta dalla seguente funzione:

$$x_j^{[s]} = F\left(\sum_{i=0}^n \sin\left(w_{ji}^{[s]} \cdot x_i^{[s-1]}\right)\right) \quad (6)$$

15

la funzione seno introduce il procedimento quali-
tativo voluto. Può essere notato come ciascun peso $w_{ji}^{[s]}$
agisce come un parametro $2\pi /$ lunghezza d'onda nella
20 coordinata i -th dello spazio degli input del nodo j -th
dello strato s -th.

Deve essere ulteriormente notato che, come già
affermato sopra, e come appare chiaramente dalla de-
scrizione su indicata, la famiglia delle reti secondo

la presente invenzione consiste in reti definite piuttosto generalmente, profondamente modificate nel comportamento interno dei loro nodi. Questa variazione dei nodi viene presa in considerazione, sia nella valutazione degli output dagli input che nelle fasi di apprendimento.

In quanto segue viene descritto con grande dettaglio un esempio di una rete neurale artificiale secondo la presente invenzione in combinazione con una ben nota struttura di una rete neurale cosiddetta Back Propagation.

La rete neurale artificiale presenta una struttura multistrato come descritto nella figura 4.

Le reti neurali Back Propagation funzionano secondo un principio ben noto cosiddetto Principio di Discesa del Gradiente (per maggiori dettagli vedere Rumelhart D.E., Hinton G. E., e Willimas R.J.) (1986), "Learning internal representations by error propagation", in Rumelhart D.E. e McClelland J.L., eds. (1986), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1, 318-362, Cambridge, MA: The MIT Press.1986).

Come esempio di una rete neurale artificiale secondo la presente invenzione la rete neurale descritta applica un principio di discesa del gradiente modificato con l'elaborazione dei nodi secondo la presente invenzione e le equazioni che governano la rete neurale saranno analizzate in dettaglio come una diretta applicazione dell'elaborazione dei nodi definita nella

equazione (5). Queste equazioni sono presentate in due gruppi: le equazioni per il trasferimento del segnale dall'input all'output (equazioni della fase in avanti) e le equazioni per la fase di apprendimento (equazioni della fase di apprendimento).

Le seguenti equazioni si riferiscono al caso generico di discesa del gradiente con multipli strati nascosti modificati secondo l'invenzione. Nella figura 4, è rappresentata una rete multistrato con due strati nascosti. Può essere notato che la presenza delle unità di ricezione che eseguono la prima sotto-fase non lineare della prima fase di trasformazione sugli input è prevista solo in corrispondenza dei nodi nascosti e di output.

Equazioni della fase in avanti

Come affermato nell'equazione (5), per lo strato(strati) generico(generici), un cosiddetto input netto armonico viene definito mediante l'equazione.

$$I_j^{[s]} = \frac{2\pi}{n} \sum_{i=0}^n \sin(w_{ji}^{[s]} \cdot x_i^{[s-1]}) \quad (7)$$

20

Il valore di attivazione negli output viene calcolato secondo l'equazione:

$$x_j^{[s]} = f(I_j^{[s]}) = f\left(\frac{2\pi}{n} \sum_i \sin(w_{ji}^{[s]} \cdot x_i^{[s-1]})\right) \quad (8)$$

25 dove, per la funzione di attivazione $f(I_j^{[s]})$, possono

essere utilizzate le usuali trasformazioni, come per esempio, la sigmoideale:

$$x_j^{[s]} = \text{sigm}(I_j^{[s]}) = \frac{1}{1 + e^{-I_j^{[s]}}} \quad (9)$$

Equazioni della fase di apprendimento

5 Viene utilizzato il principio di discesa del gradiente:

$$\Delta w_{ji}^{[s]} = -lcoef \cdot \frac{\partial E}{\partial w_{ji}^{[s]}} \quad (10)$$

con le usuali funzioni dell'Errore Globale, come
10 per esempio, l'Errore Quadratico Medio:

$$E = \frac{1}{2} \cdot \sum_{k=1}^m (t_k - x_k^{[out]})^2 \quad (11)$$

al fine di valutare l'errore su ciascun nodo viene definito un errore locale, secondo l'equazione:

$$e_j^{[s]} = -\frac{\partial E}{\partial I_j^{[s]}} \quad (12)$$

15

da quanto sopra segue:

$$\Delta w_{ji}^{[s]} = -lcoef \cdot \frac{\partial E}{\partial w_{ji}^{[s]}} =$$

$$= -lcoef \cdot \frac{\partial E}{\partial I_j^{[s]}} \cdot \frac{\partial I_j^{[s]}}{\partial w_{ji}^{[s]}} = \quad (13)$$

$$= lcoef \cdot e_j^{[s]} \cdot \frac{\partial}{\partial w_{ji}^{[s]}} \left(\frac{2\pi}{n} \sum_k \sin(w_{jk}^{[s]} \cdot x_k^{[s-1]}) \right) =$$

$$= lcoef \cdot e_j^{[s]} \cdot \frac{2\pi}{n} x_i^{[s-1]} \cdot \cos(w_{ji}^{[s]} \cdot x_i^{[s-1]})$$

5 per determinare il valore di correzione dei pesi,
 è necessario calcolare l'errore locale $e_j^{[s]}$.
 Per lo strato di output:

$$e_j^{[out]} = - \frac{\partial E}{\partial I_j^{[out]}} =$$

$$= - \frac{\partial E}{\partial x_j^{[out]}} \cdot \frac{\partial x_j^{[out]}}{\partial I_j^{[out]}} = \quad (14)$$

$$= - \left(\frac{\partial}{\partial I_j^{[out]}} f(I_j^{[out]}) \right) \cdot \left(\frac{\partial}{\partial x_j^{[out]}} \left(\frac{1}{2} \sum_{k=1}^m (t_k - x_k^{[out]})^2 \right) \right) =$$

$$= f'(I_j^{[out]}) \cdot (t_j - x_j^{[out]})$$

Per gli altri strati:

$$\begin{aligned}
e_j^{[s]} &= -\frac{\partial E}{\partial I_j^{[s]}} = \\
&= -\frac{\partial E}{\partial x_j^{[s]}} \cdot \frac{\partial x_j^{[s]}}{\partial I_j^{[s]}} = \\
&= -f'(I_j^{[s]}) \cdot \sum_k \left(\frac{\partial E}{\partial I_k^{[s+1]}} \cdot \frac{\partial I_k^{[s+1]}}{\partial x_j^{[s]}} \right) = \\
&= -f'(I_j^{[s]}) \cdot \sum_k \left(-e_k^{[s+1]} \cdot \frac{\partial}{\partial x_j^{[s]}} \left(\frac{2\pi}{n} \sum_h (\sin(w_{kh}^{[s+1]} \cdot x_h^{[s]})) \right) \right) = \\
&= f'(I_j^{[s]}) \cdot \frac{2\pi}{n} \sum_k (e_k^{[s+1]} \cdot w_{kj}^{[s+1]} \cdot \cos(w_{kj}^{[s+1]} \cdot x_j^{[s]}))
\end{aligned}$$

in quanto segue la rete neurale artificiale sopra
 10 definita viene applicata per risolvere alcuni tipici
 problemi di verifica per le reti neurali e vengono
 confrontati i risultati ottenuti alla classica rete
 neurale back propagation secondo lo stato dell'arte.

Come primo esempio viene risolto un problema mol-
 15 to semplice (il cosiddetto problema giocattolo) dalla
 rete neurale artificiale secondo l'invenzione, in modo
 da evidenziare alcune proprietà base e delle differen-
 ze da altri algoritmi.

Questo primo esempio è un ben noto semplice in-

sieme di dati noto come problema XOR. Il modello deve apprendere la regola Booleana XOR rappresentata nella seguente tabella 1

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

5

Viene fatto un confronto dei risultati ottenuti dalla rete neurale artificiale secondo la presente invenzione mediante una cosiddetta rete MLP addestrata con un algoritmo Back-propagation. La rete MLP utilizzata in questa prova è una versione migliorata del classico algoritmo Back-propagation sviluppata al Centro di Ricerca del Semeion e descritta in Buscema M. and Sacco P.L. (2000) " Feedforward networks in financial predictions: the future that modifies the present", in Expert Systems, Vol. 17(3), Agosto, Blackwell, Oxford UK. Pp.149-170, 2000.

Ovviamente entrambe le reti sono in grado di convergere alla soluzione corretta, per cui quello che è stato analizzato è la qualità della soluzione e la velocità della convergenza.

La figura 5 illustra la soluzione tipica della rete back propagation, con due iperpiani che dividono lo spazio in tre regioni. Risulta chiara la profonda

differenza tra la soluzione tipica trovata dalla rete secondo la presente invenzione illustrata nella figura 7 e la soluzione della figura 5. L'architettura della rete neurale artificiale secondo l'invenzione presenta
5 una proprietà di localizzazione simile ad una certa architettura neurale di "raggruppamento input" come la rete RBF (figura 6). Questa proprietà è dovuta alla funzione non monotona delle connessioni tra lo strato di input e quello nascosto.

10 Per quanto riguarda la velocità della convergenza, dalla tabella 2 appare evidente che la rete neurale artificiale secondo l'invenzione presenta una buona prestazione rispetto alla rete neurale back propagation. Questo è probabilmente in parte stimato dalla
15 "fase di raggruppamento" eseguita dai pesi non lineari e non monotoni.

	BP	SN
MEDIA	94.4705950.61111	
STDEV	109.328167.78612	

Il confronto della tabella 2 riguarda la velocità
20 della convergenza sul problema XOR in epoche, con 20 prove, essendo entrambe le reti fermate due volte prima della convergenza (minimo locale).

Il secondo semplice problema con cui è stato testata la rete secondo la presente invenzione è il cosiddetto "problema della spirale" il cui database è
25 illustrato graficamente nella figura 8.

Il database è stato sottoposto al calcolo con una rete back propagation con 100 unità nascoste e una rete secondo l'invenzione con 30 unità nascoste. Anche in questa prova sono stati valutati la qualità della
5 soluzione e la velocità della convergenza.

Questa volta vi è una separazione simile, con la soluzione della rete neurale secondo l'invenzione che presenta una soluzione più regolare confrontata con la separazione dello spazio ottenuta nella soluzione cal-
10 colata mediante la classica rete di back propagation.

La figura 9 illustra graficamente la soluzione del problema delle due spirali calcolata con la rete secondo la presente invenzione.

La figura 10 illustra graficamente la soluzione
15 del problema delle due spirali calcolata con la classica rete di back propagation.

Durante questa prova è stata notata una interessante proprietà dell'algoritmo secondo la presente invenzione riguardo al suo intervallo di inizializzazione
20 ne dei pesi. Da come appare evidente dalla figure 11a a 11d, esiste una irregolarità crescente della soluzione del problema della spirale a fronte di un aumentato intervallo dell'inizializzazione dei pesi, così come anche una convergenza più veloce. Questo signifi-
25 ca che è possibile controllare la compensazione tra le proprietà di approssimazione e di generalizzazione della rete secondo la presente invenzione utilizzando differenti intervalli per l'inizializzazione dei pesi.

Nella figura 11a la rete secondo l'invenzione è

stata addestrata per circa 5000 epoche con un intervallo di inizializzazione casuale tra $(-1:+1)$. Si ottiene un valore RMSE sull'insieme di addestramento di circa 0.201 e una separazione regolare tra le due spirali. Aumentando il detto intervallo di inizializzazione dei pesi ad un intervallo $(-\Pi: \Pi)$ (figura 11b) viene ottenuto lo stesso valore RMSE a 3000 epoche e la regolarità della soluzione è ancora appropriata, invece quando si utilizza un intervallo di $(-10:10)$ e di maggiore, appaiono solo alcune tracce del problema originale, ottenendo alla fine un'interpolazione punto per punto dell'insieme di addestramento.

In questo modo è possibile dire che utilizzando una rete secondo la presente invenzione è disponibile un parametro sensibile a selezionare le proprietà di interpolazione/generalizzazione del modello addestrato. Questo potrebbe essere utile, ma presenta alcuni inconvenienti se non esiste un metodo per decidere il corretto intervallo da utilizzare. Sembra che in questo caso sia stato incontrato un problema storico nella letteratura delle reti neurali, particolarmente sugli algoritmi di back propagation, definire il numero corretto di unità nascoste (vedere Geman, S., Bienenstock E. e Doursat R., (1992), "Neural Networks and the Bias/Variance Dilemma" Neural Computation, 4, 1-58, 1992).

Al contrario del numero di unità nascoste, in questo caso esiste una formula "quasi sempre vera" per questo valore ed è l'intervallo $(-\Pi: \Pi)$, che è chia-

ramente un periodo sinusoidale pieno. Poiché l'input viene normalizzato tra 0 e 1, questa scelta conduce ai migliori risultati in tutti gli esperimenti qui riferiti e in altri che confermano i risultati presentati in questa descrizione.

Un altro problema di discriminazione della spirale è la spirale con 192 punti. Questo è un problema più difficile rispetto a quello precedente. I risultati di questa prova sono illustrati nella figura 12 che illustra graficamente le capacità della rete secondo la presente invenzione di convergere su compiti difficili.

Nell'analizzare il comportamento della rete neurale artificiale secondo la presente invenzione occorre sottolineare il comportamento dal punto di vista dei pesi. Dalla semplice regola $\sin(x) \cong x$ per $x \in [0, 2]$, appare evidente che con piccoli pesi le connessioni secondo la rete della presente invenzione si trasformano in una connessione più semplice di una rete di back propagation secondo lo stato dell'arte.

È interessante osservare come la rete neurale artificiale secondo la presente invenzione utilizza la non linearità sulla connessione dei pesi durante l'apprendimento da problemi semplici a problemi più difficili. Nelle figure successive è illustrato il grafico della funzione che caratterizza ciascuna singola connessione nella rete neurale artificiale secondo l'invenzione con un intervallo di input $[0, 1]$, addendo della somma nell'equazione (5).

Nella figura 13 sono illustrati i pesi che provengono dai nodi di input ai nodi nascosti del peso del seno, in questo caso la differenza dalla semplice relazione lineare della rete di back propagation secondo lo stato dell'arte non è così forte, ma analizzando i pesi dai nodi nascosti ai nodi di output (illustrati nella figura 14) appare evidente il comportamento non lineare e non monotono della connessione sull'azione della rete neurale artificiale secondo la presente invenzione.

Nella figure 15 e 16 sono rappresentate le funzioni sui pesi da input-nascosti e nascosti-output. Aumentando la complessità del problema affrontato, il problema a 64-spirali, può essere visto che la rete neurale artificiale secondo l'invenzione utilizza una maggiore non linearità sulle connessioni accentuando in questo modo la differenza con le reti utilizzanti relazioni lineari sulle connessioni.

In ciò che segue sono analizzati il comportamento della rete neurale secondo la presente invenzione e il comportamento di una classica rete neurale di back propagation mediante una prova su due reali e complessi database.

I database di questi due ulteriori esempi sono disponibili in Internet e gli insiemi di dati selezionati sono:

- il database del cancro al seno è stato ottenuto dall'università del Ospedale del Wisconsin, Madison dal Dr. William H. Wolberg

- il database sulla valutazione del credito australiano è disponibile nel progetto Statlog.

Per effettuare le analisi su questo insieme di dati è stato scelto un approccio di validazione incrociata. Come richiesto nella documentazione, per il DB del Cancro al Seno è stata scelta una validazione incrociata per 9 volte mentre per il DB della Valutazione del Credito Australiano è stata scelta una validazione incrociata per 10 volte.

10 ~~L'utilizzo della validazione incrociata per K~~ volte è dovuto all'esplicita richiesta che si trova nella documentazione dichiarata degli insiemi di dati. Poiché questo approccio potrebbe essere inefficace per quanto riguarda l'errore del primo tipo è stato anche
15 eseguita nel confronto delle prestazioni degli algoritmi una validazione incrociata di metà campione che ha confermato i risultati riportati qui.

Il database del Cancro al seno

Le caratteristiche di questo database sono riassunte brevemente qui di seguito:

20 Titolo: Database del Cancro al Seno nel Wisconsin

Numero dei Casi: 699

Numero degli Attributi: 10 più l'attributo della classe (benigno, maligno)

25 Distribuzione della classe: Benigna: 458 (65.5%) - Maligna: 241 (34.5%)

La configurazione della rete erano per la rete back-propagation le funzioni di attivazione sigmoidale e Softmax per i nodi di output, un nodo per la classe

maligna e un nodo per quella benigna. (vedere maggiori dettagli in Mangasarian O.L. e Wolberg W.H., (1990), "Cancer Diagnosis via linear programming", SIAM News, Volume 23, Numero 5, Settembre 1990, pp 1 & 18).

5 Per la rete neurale artificiale secondo la presente invenzione è stata utilizzata la stessa configurazione.

10 Sono stati utilizzati due approcci. Il primo è stato una validazione incrociata per 9 volte con 629 osservazioni nell'insieme di addestramento e 70 osser-
vazioni nella predizione, per analizzare le capacità della rete secondo la presente invenzione per quanto riguarda alle proprietà di sovra-adattamento, ed un secondo è stato un approccio di validazione incrociata
15 per 9 volte ottenente 629 osservazioni per l'insieme di tuning e 70 osservazioni per l'insieme di predizione; l'insieme di tuning è stato diviso in insieme di addestramento (315) e insieme di verifica (314) per eseguire un approccio di "Early Stopping" per ridurre
20 il fenomeno del sovra-adattamento come descritto con maggiore dettaglio in Sarle W.S., (1995), "Stopped Training and Other Remedies for Overfitting", per apparire negli atti del 27esimo Simposio sull'Interfaccia, 1995.

25 Durante il primo esperimento sono state testate quattro differenti architetture per reti implicanti 4, 8, 12 e 16 unità nascoste al fine di valutare la sensibilità dei risultati sul numero delle unità nascoste.

La seguente tabella illustra la Media Pesata della Precisione sull'insieme di dati del Cancro al Seno

Modello		
	Nascosti Bp	SN
04	0.9525	0.9612
08	0.9481	0.9662
12	0.9519	0.9660
16	0.9543	0.9648

La seguente tabella illustra la Varianza della Precisione sull'insieme di dati del Cancro al Seno

Modello		
	HiddenBp	SN
04	0.0083	0.0072
08	0.0106	0.0070
12	0.0095	0.0078
16	0.0097	0.0076

Nelle due tabelle così come in diverse delle figure allegate la rete neurale artificiale secondo la presente invenzione è indicata come SN o rete SN o Sine Net, mentre la classica rete Back propagation secondo lo stato dell'arte è definita come Bp.

Per quanto riguarda l'assoluta prestazione la rete secondo la presente invenzione ottiene risultati migliori in tutte le configurazioni ottenendo anche una varianza inferiore degli esiti. Questo significa che la rete secondo la presente invenzione è più affi-

dabile circa il sovra-adattamento confrontata con le classiche reti back propagation. Appare anche evidente che la variabilità dei risultati è inferiore nella rete secondo la presente invenzione contro la rete back propagation. Questo fatto sottolinea che la rete secondo la presente invenzione è meno sensibile circa il numero delle unità nascoste.

Questo risultato è illustrato graficamente nella figura 17 dove è rappresentato il confronto tra una classica rete back propagation e la rete secondo la presente invenzione durante la prima prova relativamente alla correttezza del numero delle unità nascoste.

La figura 18 rappresenta il risultato del confronto tra una classica rete back propagation e la rete secondo la presente invenzione sull'insieme di dati del Cancro al Seno, confrontando le medie pesate delle unità nascoste corrette contro il numero delle unità corrette.

È stato eseguito un ulteriore test sperimentale, con 12 unità nascoste e utilizzando l'"Early Stopping". Questo ulteriore test sperimentale conferma i risultati illustrati nella figura 18. In questo caso la rete back propagation presenta un miglioramento maggiore rispetto alla rete secondo la presente invenzione sottolineando le prestazioni migliori della rete secondo l'invenzione contro il sovra-adattamento.

La figura 19 illustra analogamente alla figura 18 il risultato del confronto tra la classica rete back

propagation e la rete secondo l'invenzione
sull'insieme di dati del Cancro al Seno utilizzando
l'"Early Stopping".

Il database della Valutazione del Credito Australiano

5 Titolo: Approvazione del Credito Australiano

Numero dei Casi: 690

Numero degli Attributi: 14 + 1'attributo della classe
(buono/cattivo)

Distribuzione della Classe: Classe2: 307 (44.5%),

10 ~~Classel: 383~~ (55.5%)

Questo insieme di dati si riferisce alle richie-
ste delle carte di credito.

Sono stati utilizzati due approcci per esaminare
questo insieme di dati. Una validazione incrociata per
15 10 volte con addestramento e predizione, e una valida-
zione incrociata per 10 volte con addestramento, veri-
fica, e predizione utilizzando Early Stopping contro
il sovra-adattamento. Le configurazioni della rete so-
no le stesse dell'esperimento del Cancro al Seno.

20 I risultati della prima prova, con un numero di
unità nascoste di 4, 8 , 12, 16 sono riassunti nella
figura 20 e nella seguente tabella 5.

VI 10 volte		
Nasco-		
sti	Bp	SN
04	0.8146	0.8573
08	0.8333	0.8449
12	0.7981	0.8512

16 0.8220 0.8525

La tabella 5 illustra il confronto tra una classica rete back propagation Bp e la rete secondo l'invenzione chiamata rete SN durante la prima prova e
5 collegato alla correttezza del numero delle unità nascoste.

La figura 20 è una rappresentazione grafica dei risultati della tabella 5.

Similmente al precedente test sperimentale nel
10 secondo test sono state utilizzate 12 unità nascoste, e il confronto tra i risultati della prima prova è riportato nella tabella 6 e nella figura 21.

	Ad-Pr		Ad-Vr-Pr	
	BP	SN	BP	SN
Media	0.8146	0.8573	0.8629	0.8653
Var	0.0022	0.0018	0.0033	0.0024

15 La tabella 6 riassume il risultato del confronto tra una classica rete back propagation Bp e la rete secondo la presente invenzione SN sull'insieme di dati della Valutazione del Credito Australiano, con "Early Stopping" (Addestramento-Verifica-Predizione) e senza
20 (Addestramento-Predizione). La figura 21 illustra una rappresentazione grafica dei risultati della tabella 6.

I risultati assoluti e il confronto tra la prima e la seconda prova confermano gli esiti ottenuti con

l'insieme di dati del Cancro al Seno: vale a dire risultati migliori della rete secondo l'invenzione con una varianza più piccola e una maggiore differenza della correttezza tra la prima e la seconda prova per
5 la rete back propagation, suggerendo il comportamento migliore della rete secondo l'invenzione circa il problema del sovra-adattamento.

RIVENDICAZIONI

1. Una rete neurale, comprendente diversi nodi che formano almeno due strati uno dei quali è il cosiddetto strato di input e il secondo dei quali è il
5 cosidetto strato di output, i nodi dello strato di input formando degli input per l'immissione di dati di input di un database e i nodi dello strato di output formando dei canali di output per fornire i dati di
10 output che sono i risultati dell'elaborazione dei dati di input, essendo previste delle connessioni tra i nodi dello strato di input e i nodi dello strato di output, ciascun nodo dello strato di output eseguendo una trasformazione dei dati di input ricevuti dallo
15 strato di input in dati di output la quale trasformazione comprende una prima fase di trasformazione consistente essenzialmente nel sommare i dati di input ricevuti dai nodi di input ai detti nodi di output pesando i detti dati di input e una seconda fase di trasformazione che trasforma, non linearmente, i risultati
20 ottenuti mediante la prima fase di trasformazione, i dati di output ottenuti mediante detta trasformazione eseguita in un nodo di output essendo i dati di output
25 caratterizzato dal fatto che
in ciascun nodo di output la prima fase di trasformazione è una funzione di trasformazione non lineare dei dati di input ricevuti mediante i nodi di output dai nodi di input.

2. Una rete neurale artificiale secondo la rivendicazione 1 comprendente uno strato di input avente un numero predeterminato di nodi di input e uno strato di output avente un numero predeterminato di nodi di output, caratterizzata dal fatto che tra lo strato di input e quello di output è previsto almeno un ulteriore strato di nodi, il cosiddetto strato nascosto, o più di uno strato nascosto, i nodi di questo strato nascosto essendo connessi mediante connessioni pesate ai nodi di input dello strato di input e ai nodi di un ulteriore strato nascosto quando è previsto più di uno strato nascosto o ai nodi di output dello strato di output se è previsto solo uno strato nascosto; ciascun nodo dell' almeno uno strato nascosto o di più di uno degli strati nascosti e i nodi dello strato di output eseguendo una trasformazione dei dati di input ricevuti dallo strato di input o da un precedente strato nascosto in dati di output, la quale trasformazione comprende una prima fase di trasformazione consistente nel sommare i dati di input ricevuti dai nodi di input ai detti nodi di output pesando i detti dati di input e una seconda fase di trasformazione che trasforma in modo non lineare i risultati ottenuti dalla prima fase di trasformazione, i dati di output ottenuti da questa trasformazione eseguita in detti nodi essendo i dati di output se i nodi sono i nodi di output dello strato di output o i nodi di input forniti dai nodi di uno strato nascosto ai nodi di un successivo strato nascosto o ai nodi di output dello strato di output, in

ciascun nodo di ciascun strato nascosto e nei nodi di output dello strato di output la prima fase di trasformazione è una funzione di trasformazione non lineare dei dati di input ricevuti mediante i nodi di output o mediante i nodi di uno strato nascosto dai nodi di input dello strato di input o mediante i nodi del precedente strato nascosto.

3. Una rete neurale artificiale secondo la rivendicazione 1 in cui i dati di input dei nodi dello strato di input consistono nei dati di input del database, mentre i dati di output dei nodi dello strato di input sono forniti ai nodi dello strato di output o ai nodi del primo strato nascosto o ai nodi dell'almeno uno strato nascosto come dati di input dei nodi di questi strati e i dati di output dello strato di output consistono nel risultato dell'elaborazione della rete neurale artificiale.

4. Una rete neurale artificiale in cui la prima trasformazione non lineare dei dati di input eseguita da un nodo comprende una prima sotto-fase di trasformazione dei dati di input mediante una funzione sinusoidale e una seconda sotto-fase di trasformazione consistente nella somma dei dati di input dopo la trasformazione mediante la funzione sinusoidale, cioè dopo aver eseguito la prima sotto-fase di trasformazione.

5. Una rete neurale artificiale secondo una o più delle precedenti rivendicazioni, caratterizzata dal fatto che ciascun nodo almeno dell'almeno uno strato

nascosto e dello strato di output comprende diversi canali di input per differenti dati di input;

ciascun canale essendo associato

un'unità di ricezione per l'esecuzione della prima sotto-fase di trasformazione non lineare della prima fase di trasformazione;

un'unità di sommatoria per l'esecuzione della seconda sotto-fase di trasformazione della prima fase di trasformazione sommando i dati di input di ciascun canale trasformati non linearmente ad un valore

e un'unità di trasformazione non lineare per l'esecuzione della seconda fase di trasformazione mediante un filtraggio non lineare del valore ottenuto mediante la prima fase di trasformazione e per fornire il valore di output del nodo che è il valore di input dei nodi di un successivo strato nascosto o dello strato di output.

6. Una rete neurale artificiale secondo una o più delle precedenti rivendicazioni, caratterizzata dal fatto che i dati di input consistono in un numero predeterminato di variabili in uno spazio di variabili dei dati di input ciascuna variabile essendo definita da coordinate nello spazio dei dati di input e ciascuna coordinata nello spazio dei dati di input è trasformata non linearmente nella prima fase di trasformazione in un corrispondente valore di una variabile che viene reso dipendente dalla posizione nello spazio del valore della coordinata rispetto un'onda spaziale di lunghezza d'onda data, questa dipendenza consisten-

te nel moltiplicare i valori delle coordinate di input per la lunghezza d'onda di un'onda sinusoidale che vengono poi trasformate nello stesso valore, la lunghezza su ciascuna coordinata di input essendo regolata durante la fase di apprendimento.

7. Una rete neurale artificiale caratterizzata dal fatto che la trasformazione dei dati di input eseguita da ciascun nodo viene definita dalla seguente equazione:

$$x_j^{[s]} = F\left(G\left(w_{ji}^{[s]}, x_i^{[s-1]}\right)\right) \quad (4)$$

dove la trasformazione non lineare $F(\cdot)$ è la funzione di filtraggio non lineare della seconda fase di trasformazione e $G(\cdot)$ è la funzione di trasformazione non lineare della prima fase di trasformazione consistente nella somma degli input pesati elaborati in modo non monotono, sinusoidale secondo la seguente funzione:

$$G\left(w_{ji}^{[s]}, x_i^{[s-1]}\right) = \sum_{i=0}^n \sin\left(w_{ji}^{[s]} \cdot x_i^{[s-1]}\right) \quad (5)$$

dove

$[s]$: lo strato generico della rete, con $s=1$ per lo strato di input e valori crescenti per gli strati nascosti e di output;

$x_j^{[s]}$: la variabile di output del nodo j -th nello strato $[s]$;

$x_i^{[s-1]}$: l'input i -th del nodo generico nello strato $[s]$ dal nodo i -th nello strato $[s-1]$;

5 $x_0^{[s-1]}$: un input "falso" al nodo generico nello strato $[s]$, introdotto artificialmente per rappresentare, in un modo matematicamente conveniente, un utile valore di soglia che solitamente è fissato a 1.

10 $w_{ji}^{[s]}$: il peso sulla connessione che unisce il nodo i -th nello strato $[s-1]$ al nodo j -th nello strato $[s]$;

n : il numero di input al nodo.

8. Una rete neurale artificiale secondo la rivendicazione 7, caratterizzata dal fatto che ciascun nodo
15 esegue una trasformazione dei dati di input secondo la seguente funzione:

$$x_j^{[s]} = F \left(\sum_{i=0}^n \sin(w_{ji}^{[s]} \cdot x_i^{[s-1]}) \right) \quad (6)$$

La funzione del seno introducendo un processo qualitativo poiché ciascun peso $w_{ji}^{[s]}$ agisce come un

parametro 2π / lunghezza d'onda nella coordinata i -th dello spazio degli input del nodo j -th dello strato s -th.

5 9. Una rete neurale artificiale secondo una o più delle precedenti rivendicazioni caratterizzata dal fatto che la seconda fase di trasformazione non lineare è eseguita mediante una funzione sigmoideale.

10 10. Una rete neurale artificiale secondo una o più delle precedenti rivendicazioni caratterizzata dal fatto che essa è una rete neurale Back propagation multistrato comprendente una fase in avanti e una fase di apprendimento che utilizza un principio di discesa del gradiente;

15 la fase in avanti essendo definita dalle seguenti equazioni:

una prima fase di trasformazione armonica che fornisce il valore I dell'input netto trasformato dove

$$I_j^{[s]} = \frac{2\pi}{n} \sum_{i=0}^n \sin(w_{ji}^{[s]} \cdot x_i^{[s-1]}) \quad (7)$$

20 Una seconda fase di trasformazione non lineare mediante una cosiddetta funzione di attivazione $f(I_j^{[s]})$ determinante l'output del nodo secondo l'equazione:

$$x_j^{[s]} = f(I_j^{[s]}) = f\left(\frac{2\pi}{n} \sum_i \sin(w_{ji}^{[s]} \cdot x_i^{[s-1]})\right) \quad (8)$$

la fase di apprendimento essendo definita dalle seguenti equazioni:

25 Il principio di discesa del gradiente

$$\Delta w_{ji}^{[s]} = -lcoef \cdot \frac{\partial E}{\partial w_{ji}^{[s]}} \quad (10)$$

con le solite funzioni di Errore globale, l'errore si calcola per ciascun nodo valutando definendo il locale secondo l'equazione:

$$e_j^{[s]} = -\frac{\partial E}{\partial I_j^{[s]}} \quad (12)$$

5

Ottenendo così

$$\begin{aligned} \Delta w_{ji}^{[s]} &= -lcoef \cdot \frac{\partial E}{\partial w_{ji}^{[s]}} = \\ &= -lcoef \cdot \frac{\partial E}{\partial I_j^{[s]}} \cdot \frac{\partial I_j^{[s]}}{\partial w_{ji}^{[s]}} = \end{aligned} \quad (13)$$

$$\begin{aligned} &= lcoef \cdot e_j^{[s]} \cdot \frac{\partial}{\partial w_{ji}^{[s]}} \left(\frac{2\pi}{n} \sum_k \sin(w_{jk}^{[s]} \cdot x_k^{[s-1]}) \right) = \\ &= lcoef \cdot e_j^{[s]} \cdot \frac{2\pi}{n} x_i^{[s-1]} \cdot \cos(w_{ji}^{[s]} \cdot x_i^{[s-1]}) \end{aligned}$$

10

e per determinare il valore di correzione dei pesi, l'errore locale $e_j^{[s]}$ essendo calcolato come:

15

Per lo strato di output:

$$\begin{aligned}
e_j^{[out]} &= -\frac{\partial E}{\partial I_j^{[out]}} = \\
&= -\frac{\partial E}{\partial x_j^{[out]}} \cdot \frac{\partial x_j^{[out]}}{\partial I_j^{[out]}} =
\end{aligned}
\tag{14}$$

Per gli altri strati:

$$\begin{aligned}
e_j^{[s]} &= -\frac{\partial E}{\partial I_j^{[s]}} = \\
&= -\frac{\partial E}{\partial x_j^{[s]}} \cdot \frac{\partial x_j^{[s]}}{\partial I_j^{[s]}} = \\
&= -f'(I_j^{[s]}) \cdot \sum_k \left(\frac{\partial E}{\partial I_k^{[s+1]}} \cdot \frac{\partial I_k^{[s+1]}}{\partial x_j^{[s]}} \right) =
\end{aligned}
\tag{15}$$

11. Una rete neurale artificiale secondo la rivendicazione 10, caratterizzata dal fatto che

la funzione di attivazione della fase in avanti $f(I_j^{[s]})$ è la funzione sigmoideale:

$$x_j^{[s]} = \text{sigm}(I_j^{[s]}) = \frac{1}{1 + e^{-I_j^{[s]}}}
\tag{9}$$

la funzione di errore della fase di apprendimento è la funzione dell'Errore Quadratico Medio:

$$E = \frac{1}{2} \cdot \sum_{k=1}^m (t_k - x_k^{[out]})^2 \quad (11)$$

il valore di correzione del peso per lo strato di output diventa

$$\begin{aligned} 5 \quad & - \left(\frac{\partial}{\partial I_j^{[out]}} f(I_j^{[out]}) \right) \cdot \left(\frac{\partial}{\partial x_j^{[out]}} \left(\frac{1}{2} \sum_{k=1}^m (t_k - x_k^{[out]})^2 \right) \right) = \\ & = f'(I_j^{[out]}) \cdot (t_j - x_j^{[out]}) \end{aligned}$$

il valore di correzione del peso per gli altri strati diventa

10

$$\begin{aligned} & - f'(I_j^{[s]}) \cdot \sum_k \left(-e_k^{[s+1]} \cdot \frac{\partial}{\partial x_j^{[s]}} \left(\frac{2\pi}{n} \sum_h (\sin(w_{kh}^{[s+1]} \cdot x_h^{[s]})) \right) \right) = \\ & = f'(I_j^{[s]}) \cdot \frac{2\pi}{n} \sum_k (e_k^{[s+1]} \cdot w_{kj}^{[s+1]} \cdot \cos(w_{kj}^{[s+1]} \cdot x_j^{[s]})) \end{aligned}$$

15

12. Una rete neurale artificiale secondo una o più delle precedenti rivendicazioni, caratterizzato dal fatto che i nodi di input dello strato di input eseguono una prima fase di trasformazione non lineare e una seconda fase di trasformazione non lineare dei dati di input ricevuti essendo provvisti delle seguenti unità:

20

un canale di ricezione per i dati di input

un'unità di ricezione associata a detto canale di ricezione per eseguire la prima sotto fase di trasformazione non lineare della prima fase di trasformazione come definita nelle precedenti rivendicazioni;

5 un'unità di sommatoria per eseguire la seconda sotto fase di trasformazione della prima fase di trasformazione sommando i dati di input trasformati in modo non lineare di ciascun canale ad un valore come definito nelle precedenti rivendicazioni;

10 e un'unità di trasformazione non lineare per eseguire la seconda fase di trasformazione mediante un filtraggio non lineare del valore ottenuto mediante la prima fase di trasformazione come definita dalla precedente fase e fornendo il valore di output dei nodi
15 di input che è il valore di input dei nodi di uno strato nascosto successivo o dello strato di output.

RIASSUNTO

Una rete neurale artificiale comprende almeno uno strato di input con un numero predeterminato di nodi di input e almeno uno strato di output con un numero predeterminato di nodi di output o anche almeno uno strato nascosto intermedio con un numero predeterminato di nodi tra lo strato di input e di output. Almeno i nodi dello strato di output e/o dello strato nascosto e/o anche dello strato di input eseguono una trasformazione non lineare dei dati di input per calcolare un valore di output che deve essere fornito come un valore di input ad uno strato successivo o un dato di output se viene considerato lo strato di output.

(fig.2)

15

1/12

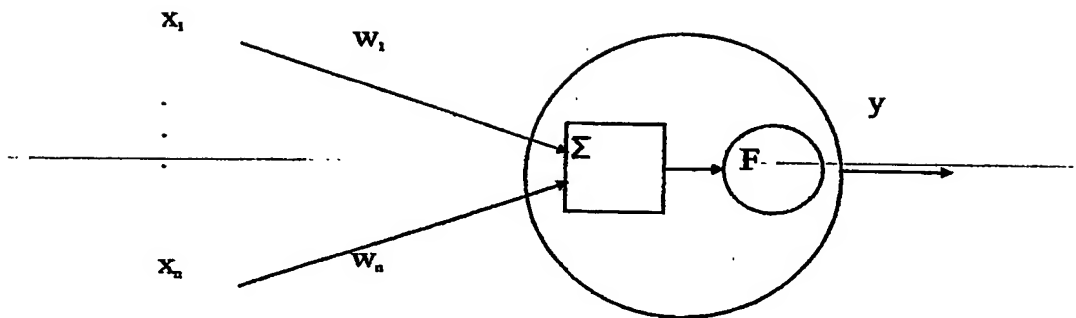


Figure 1 — Conceptual processing in classical networks

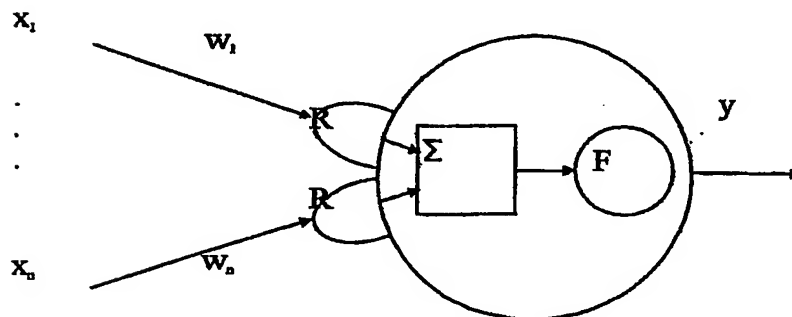


Figure 2 - Conceptual processing in classical Sine Net networks

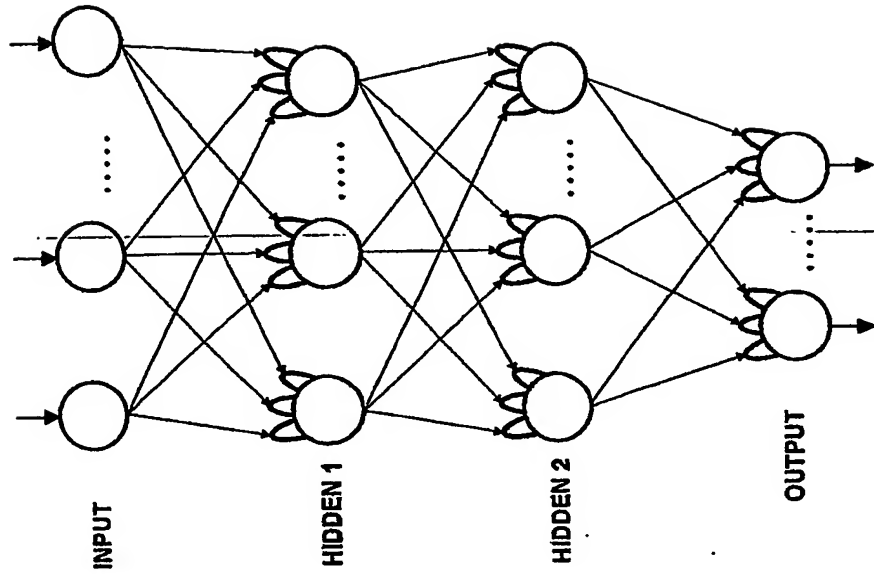


Figure 4 — Representation of a multi-layer Sine Net

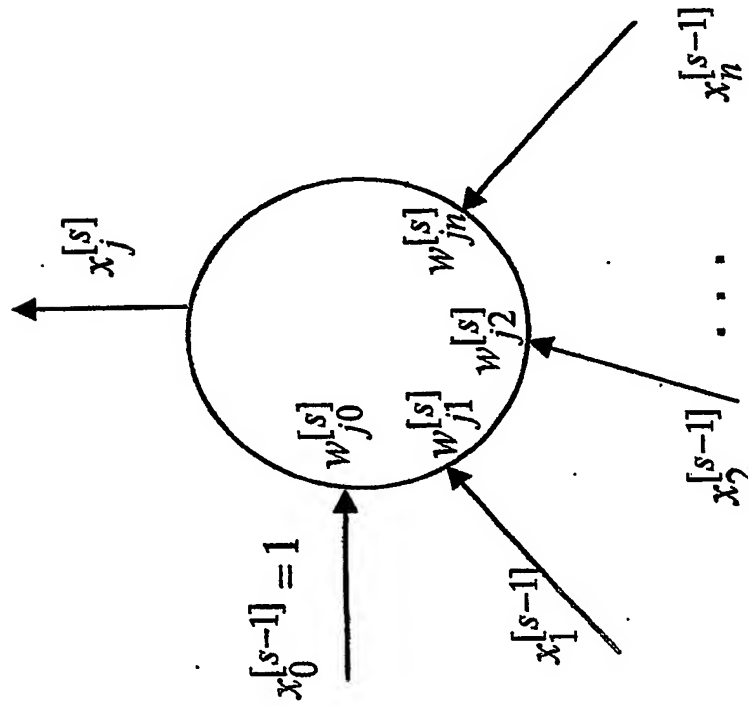


Figure 3 — Node variables for the j -th node

Figure 10 — Back-propagation solution of the two spirals problem.

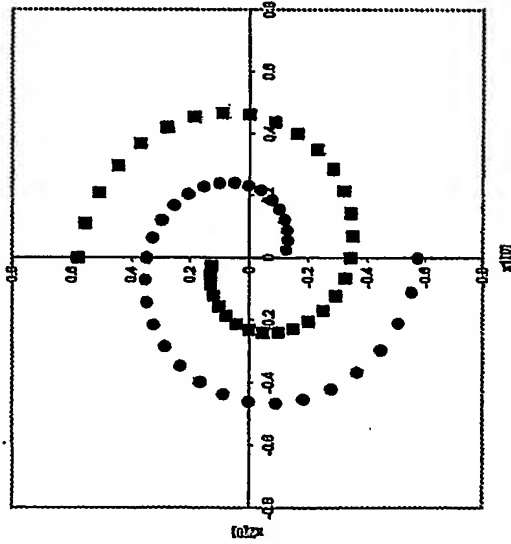


Figure 8 — The two spirals problem.

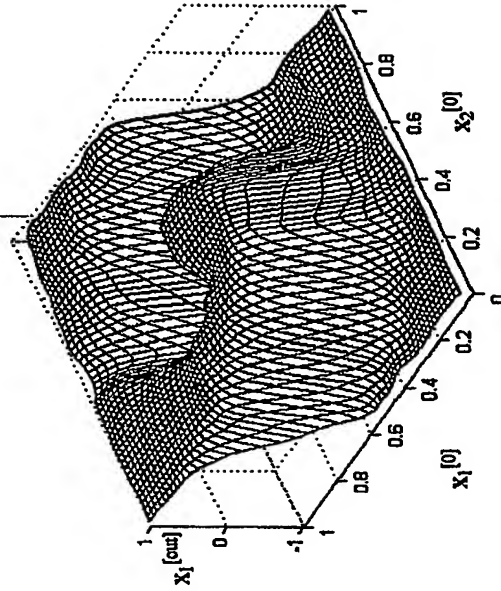
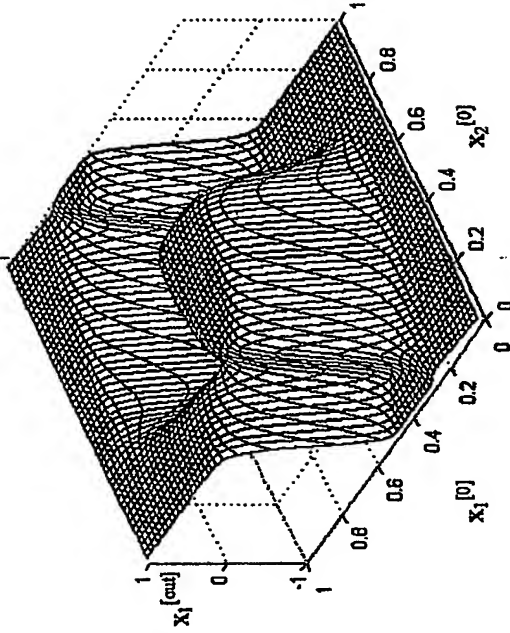


Figure 9 — Sine Network solution of the two spirals problem.

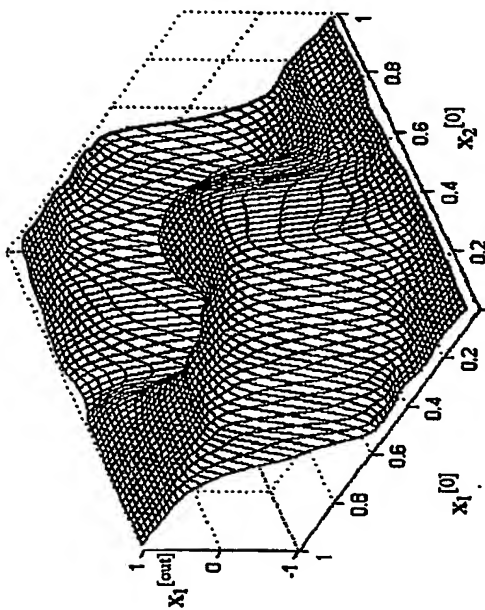


Fig. 11a – SN solution of Spiral problem with weight range $(-1,1)$

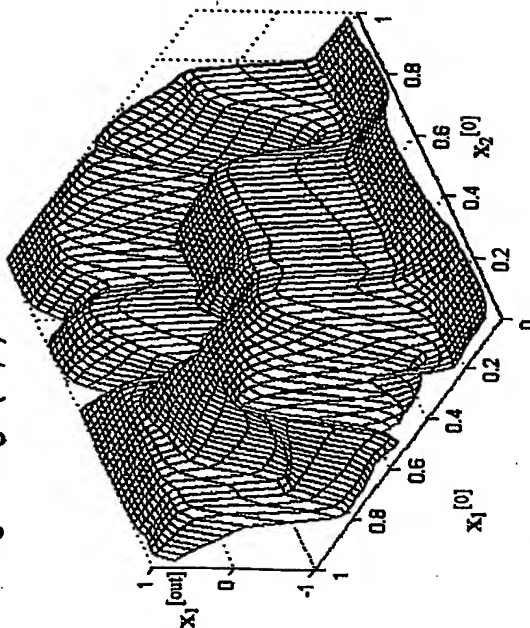


Figure 11c – SN solution of Spiral problem with weight range $(-10,10)$

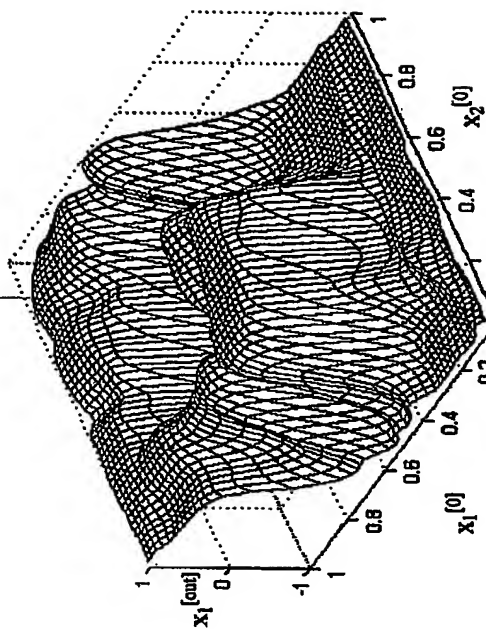


Fig. 11b – SN solution of Spiral problem with weights range $(-\pi, \pi)$

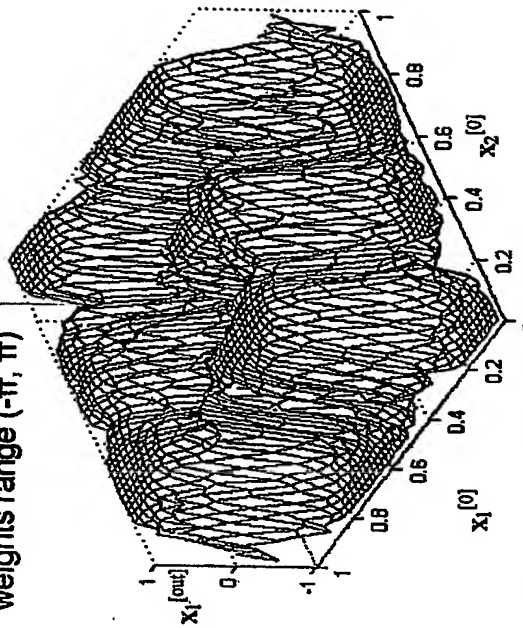


Fig. 11d – SN solution of Spiral problem with weight range $(-20,20)$

6/12

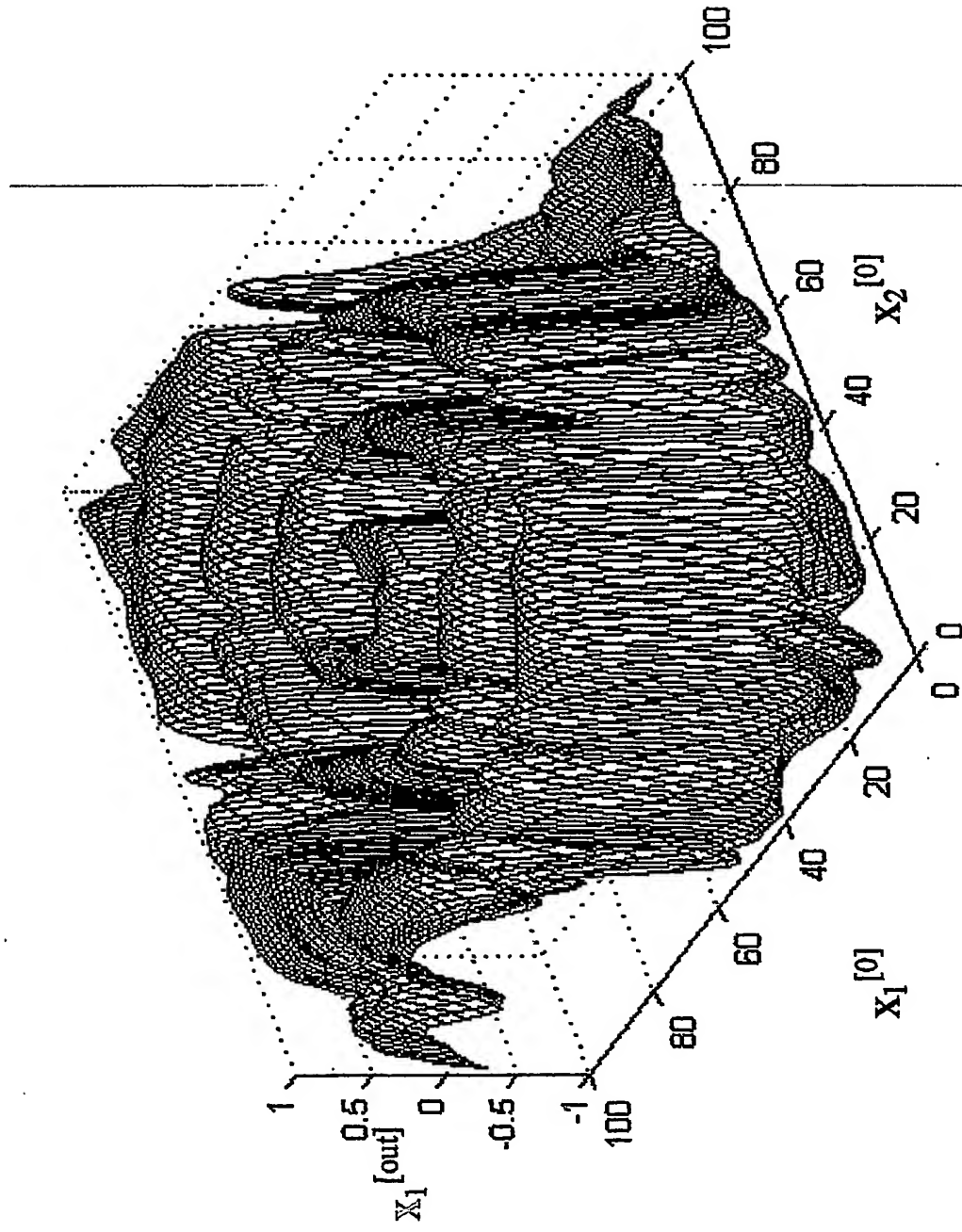


Figure 12 — Solution of Sine Net on spiral problem with 192 points.

7/12

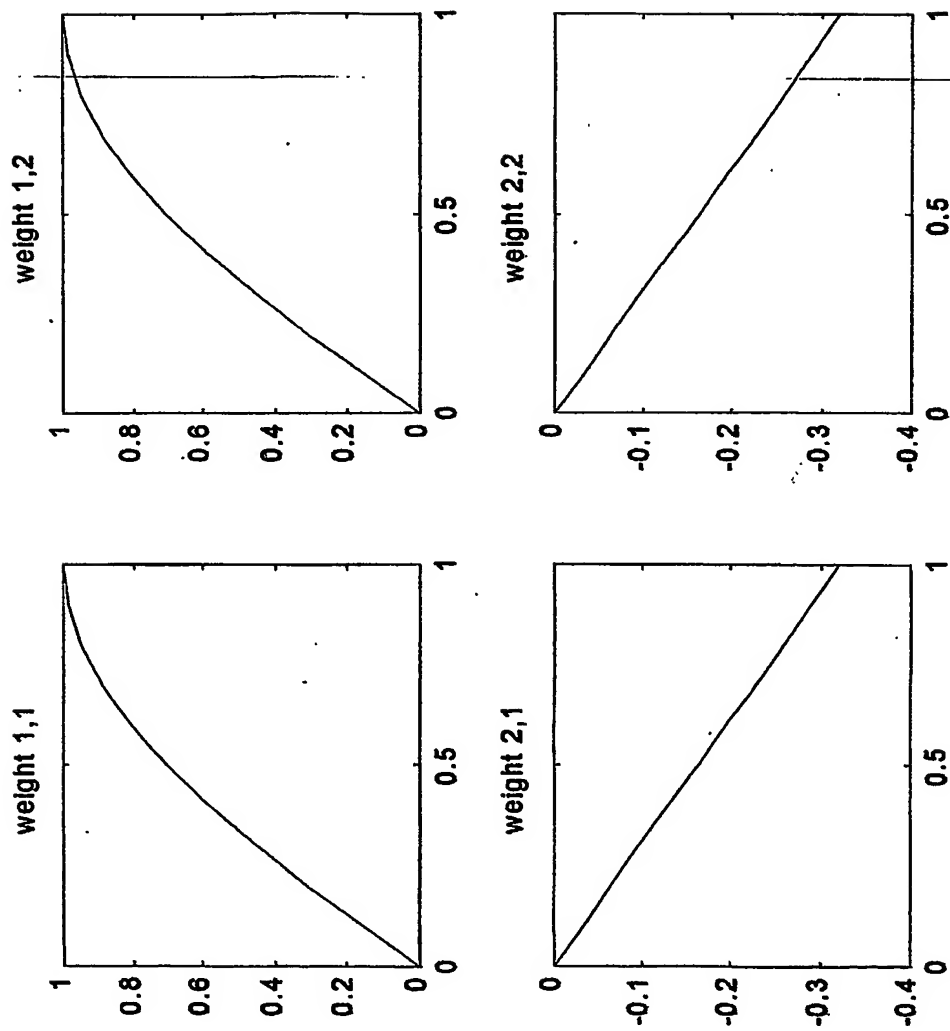


Figure 13 – SN weight transfer function from input to hidden nodes

8/12

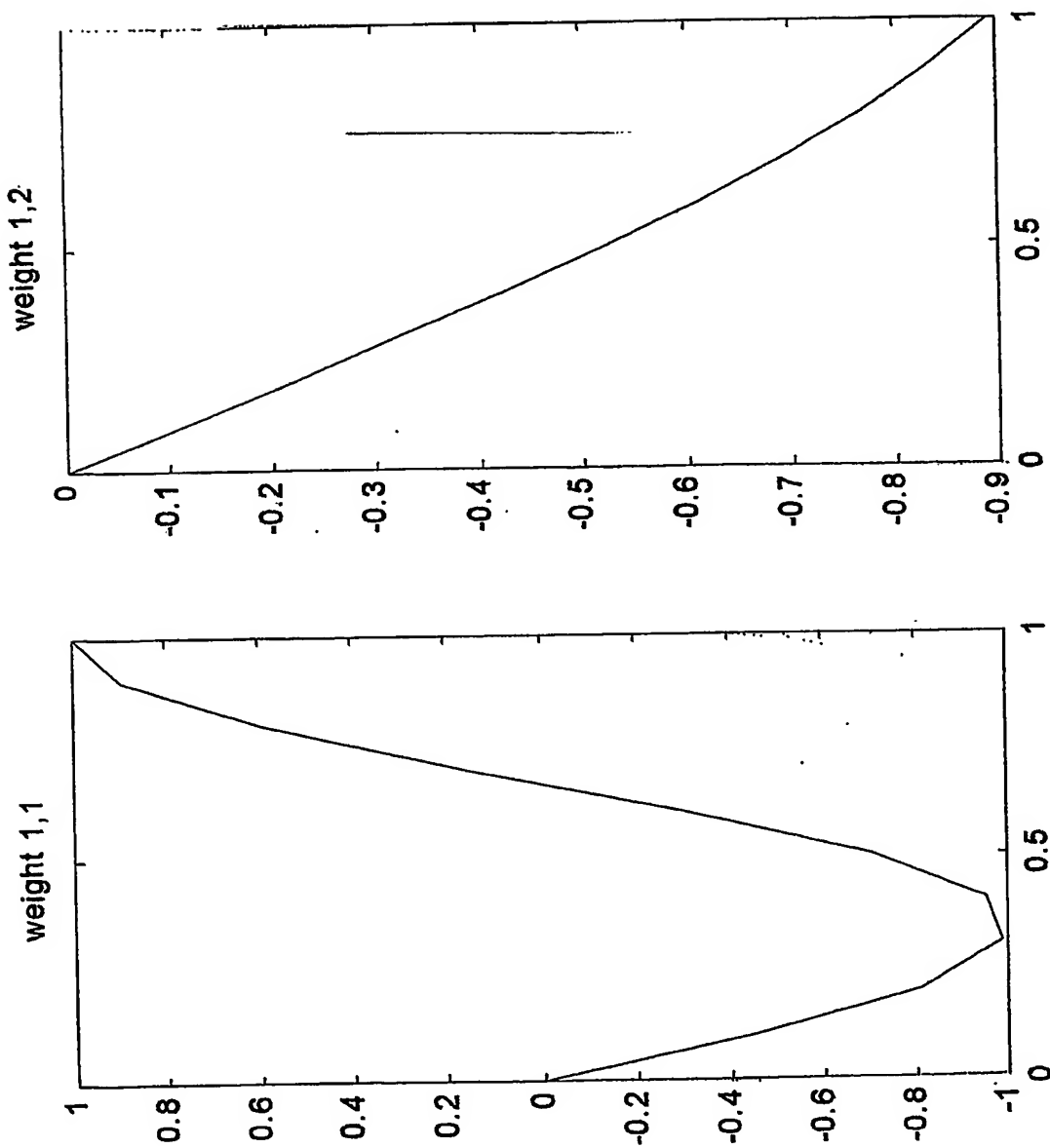


Figure 14 – SN_weight transfer function from hidden to output nodes

9/12

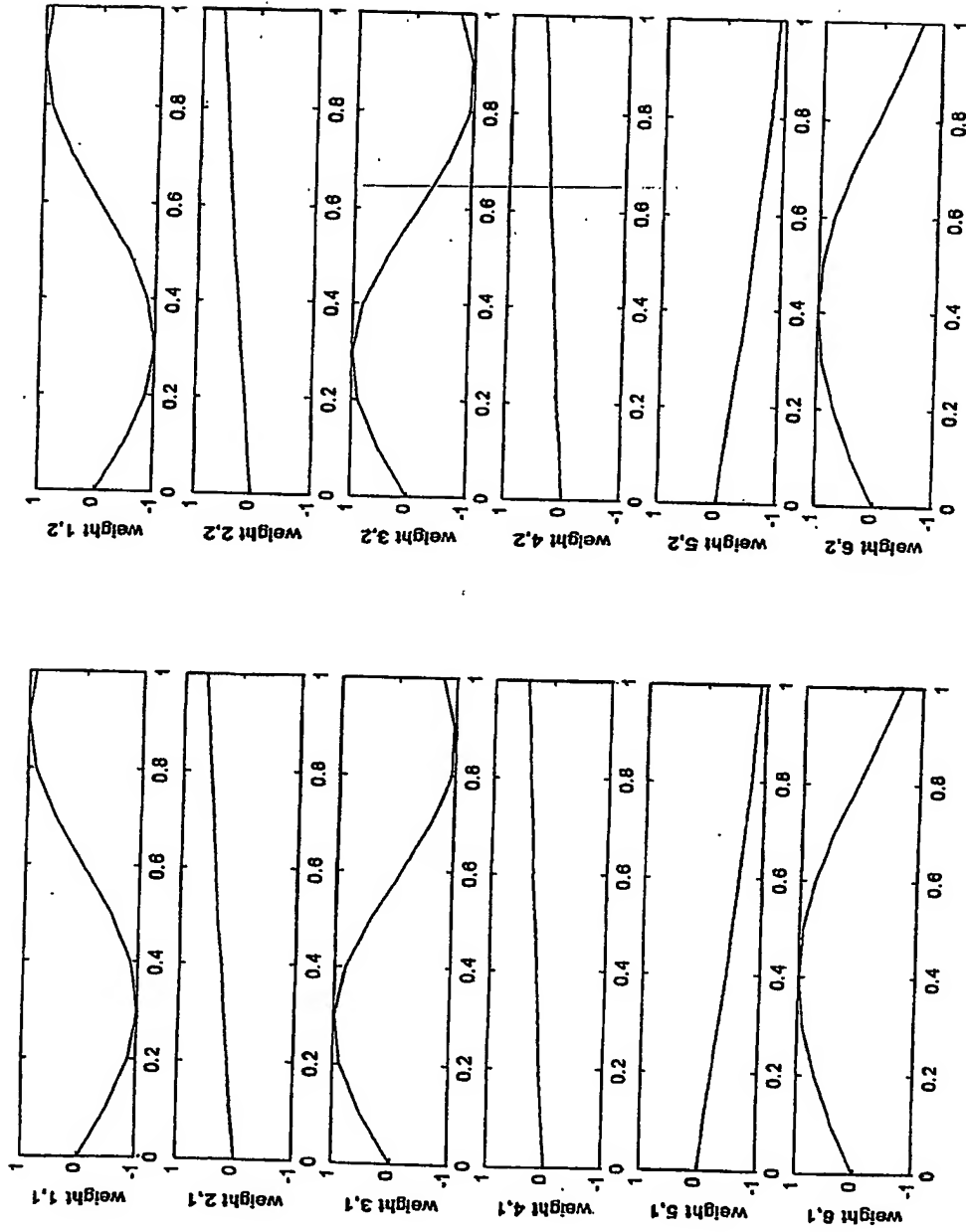


Figure 15 – SN weight transfer function from input to hidden nodes

10/12

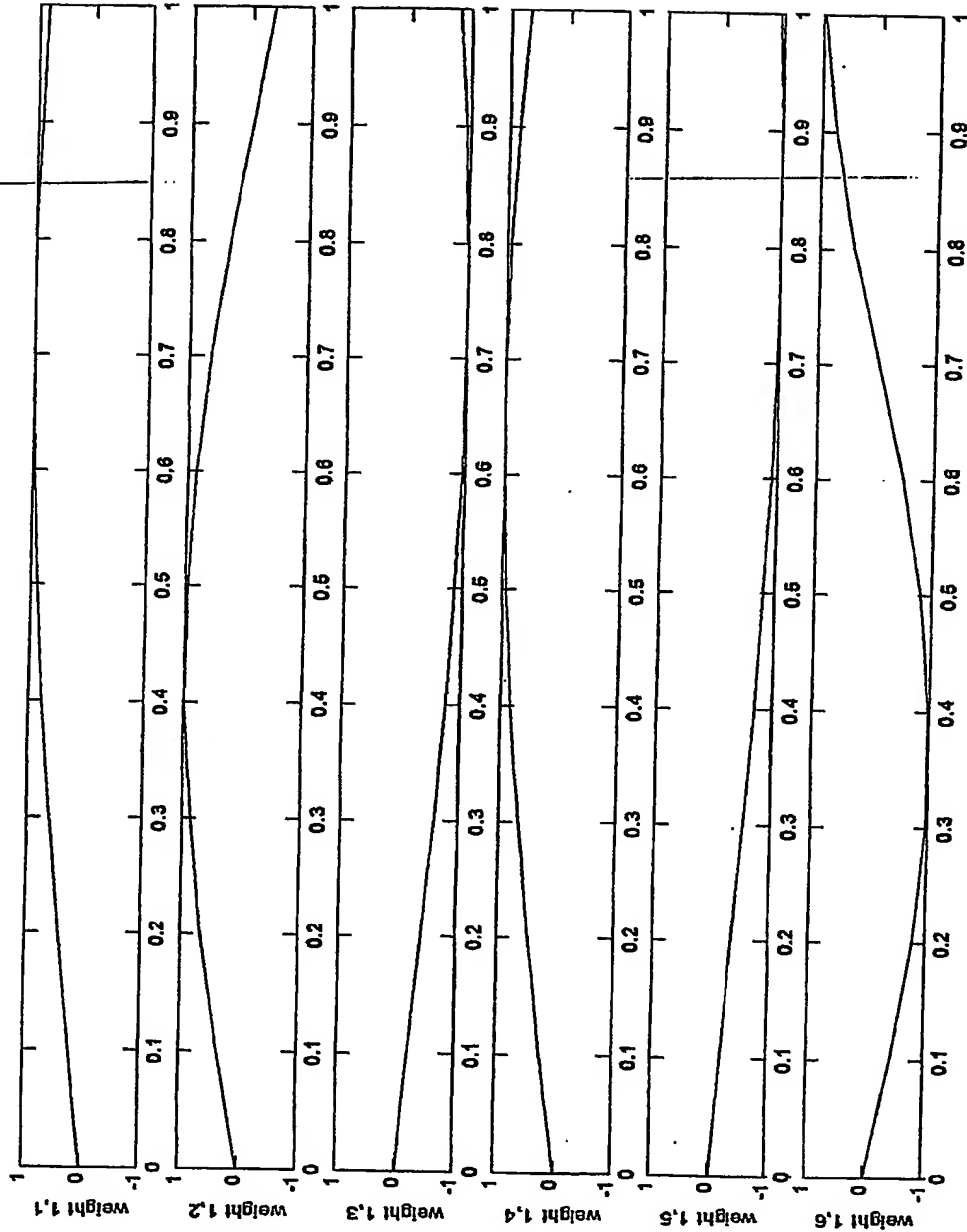


Figure 16 – SN weight transfer function from hidden to output nodes

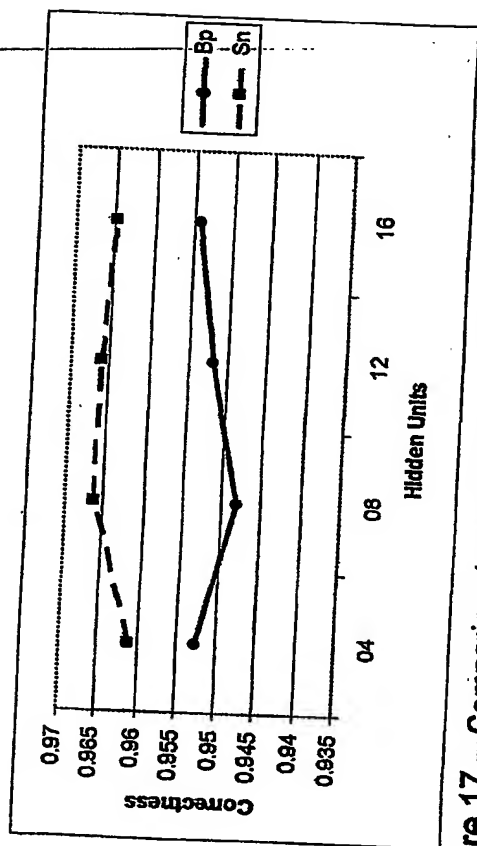


Figure 17 – Comparison between Bp and Sn network during the first trial, correctness on hidden units number

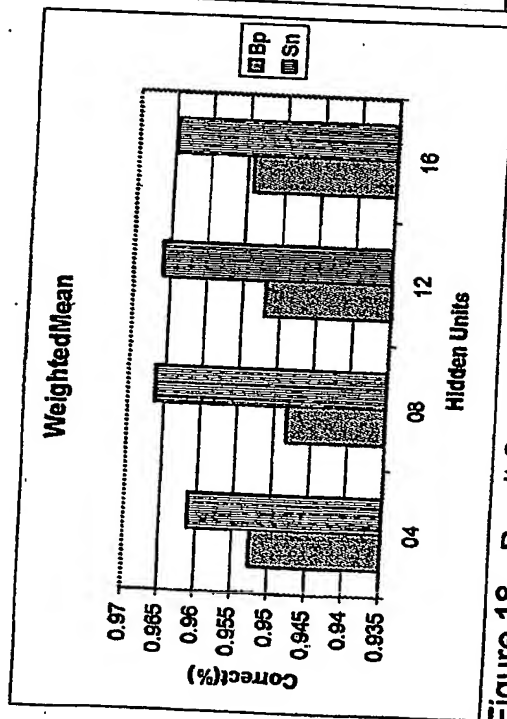


Figure 18 - Result Comparison between BP and SN on the Breast Cancer dataset.

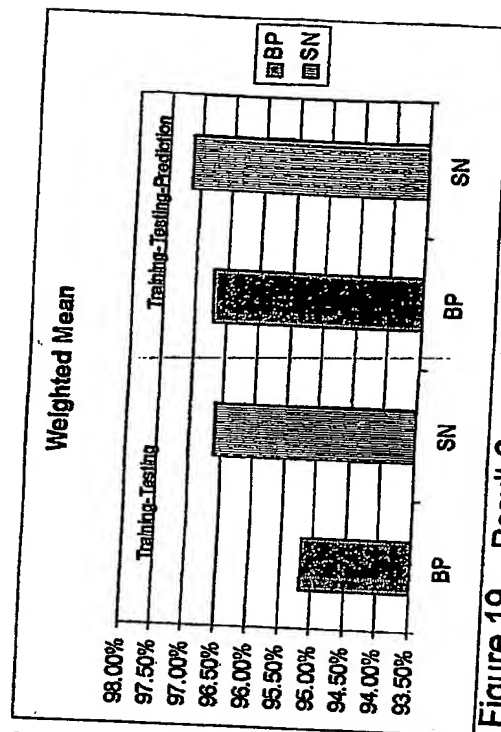


Figure 19 – Result Comparison between BP and SN on the Breast Cancer dataset using "Early Stopping"

12/12

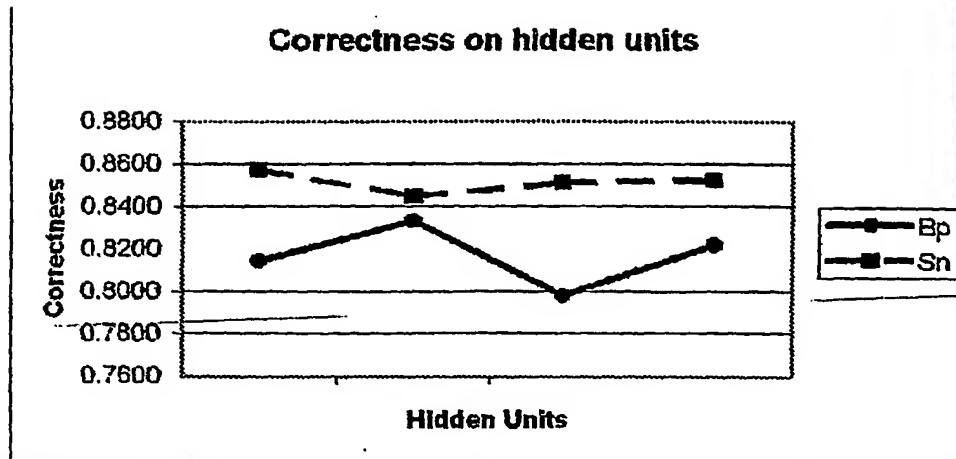


Figure 20 – Comparison between Bp and SN network during the first trial, correctness on hidden units number

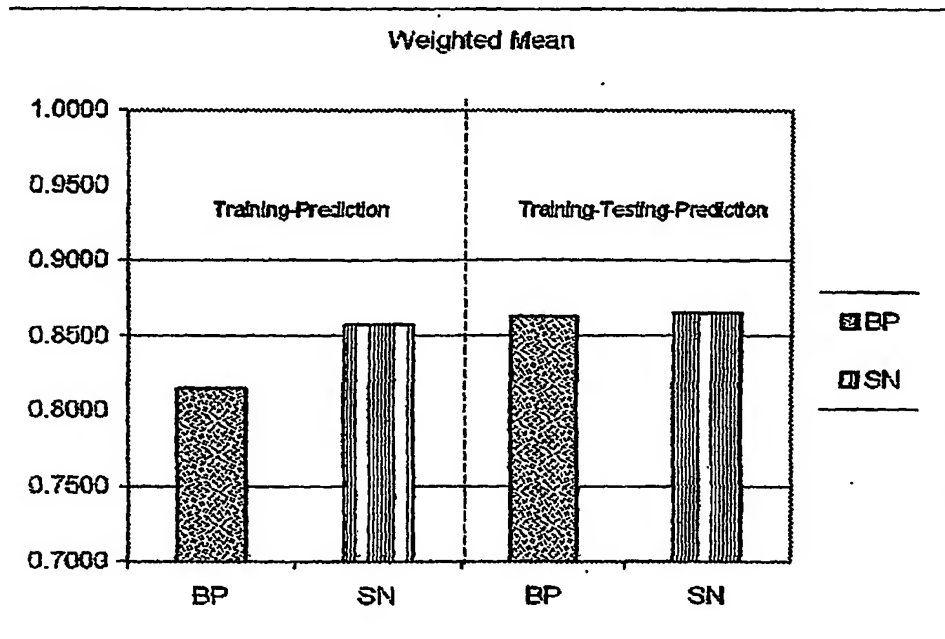


Figure 21 – Result Comparison between BP and SN on the Australian Credit Scoring dataset, with "Early Stopping" (Tr-Ts-Pr) and without (Tr-Pr)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.